ACAT-G: An Interactive Learning Framework for Assisted Response Generation

Xueyuan Lu, [1,2] **Saurav Sahay** [1]**, Zhou Yu** [3]**, Lama Nachman** [1]
[1] Intel Labs, Anticipatory Computing Lab, USA
[2] University of California, Davis
[3] Columbia University
{xueyuan.lu, saurav.sahay, lama.nachman} @intel.com, joyu@ucdavis.edu

Abstract

In this paper, we introduce ACAT-G, an interactive dialogue learning framework that incorporates constant human feedback into fine-tuning language models in order to assist conditioned dialog generation. The system takes in a limited amount of input from a human and generates personalized response corresponding to the context of the conversation within natural dialog time-frame. By combining inspirations from online learning, reinforcement learning, and large scale language models, we expect this project to provide a foundation for human-in-the-loop conditional dialog generation tasks.

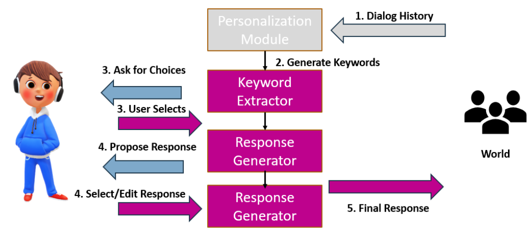

Figure 1: Overview of ACAT-G Framework. Response Generator is displayed twice, since it is called twice.

Introduction

ACAT (Assistive Context Aware Toolkit)(Nachman, Prasad et al. 2018) project is an open source platform developed at Intel Labs to enable people with motor neuron diseases to have full access to the capabilities and applications of their computers using very limited user input (e.g. gaze, single muscle movement, facial gesture, etc). In addition to providing access to applications, since many users with MND lose their ability to speak, ACAT also enables users to communicate with others by converting their text input into speech using a TTS system. ACAT utilizes word prediction to reduce the effort needed by the user to enter the text. Due to the recent advancements of large scale language models such as OpenGPT (Radford et al. 2018) and the more recent successors, next generation of ACAT aims to provide the user a high quality near real time conversational experience while minimizing the number of input required and maximizing the trust between the user and the system. We envision the new version of ACAT to listen to the speaker, convert it to text, provide different responses to the user to choose from, hence dramatically reducing the silence gap in the conversation resulting from users input their responses from scratch.

While we have several methods available to train conditional language generation models today, most of them rely on optimizing the MLE loss. These models inherently suffer from the problem of exposure bias, repetitions, and lack of diverse generation, other methods have emerged to tackle these issues using methods like unlikelihood training(Welleck et al. 2020), sequence decoding with various loss functions(See et al. 2019; Gupta et al. 2019), ranking responses(Gao et al. 2020), etc. Very few methods today apply online learning approaches to train these generative models(yang Wu, Li, and Yu 2020). In this work, we are exploring methods inspired from the TextGAIL approach(Welleck et al. 2020) to train models with reinforcement learning based online updates.

For the user to trust an assistive computing system like ACAT, the user must be able to control the output of the system. Thus, the system, in addition to providing a feedback interface in the reinforcement learning loop, must also provide the ability to control information.

To address the above questions, this paper introduces ACAT-G dialog based interaction and learning framework. Built on top of ParlAI (Miller et al. 2017) dialog collection and training pipeline, the key contribution to ACAT-G is it utilizes reinforcement learning paradigm to combine the process of training a language model with human control and feedback. We guide the development with three principles: 1. minimized user input interface, 2. online learning paradigm that can tune generated response 3. maximized sense of control over the dialog flow. This system proposes personalization, keyword extraction, and response generation modules. Personalization module provides an interface for the trained system to insert additional information about the user. Personalization module outputs a control vector that encodes the user's personal information to the keyword extractor, which propose a list of keywords relevant to the conversation.

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

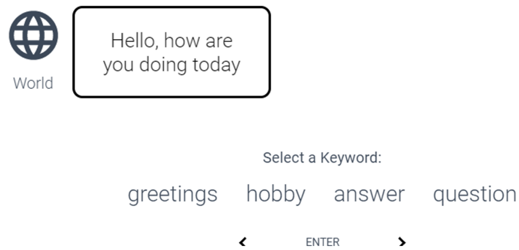

Figure 2: ACAT-G UI Keyword Selection/Editing

Figure 3: ACAT-G UI Response Generation/Editing

System

ACAT-G Framework is built on top of ParlAI. ParlAI (Miller et al. 2017) is a platform for training and evaluating dialog models with a library of formatted dialog data and datasets. ACAT-G utilizes the Agent-Teacher paradigm in ParlAI to encapsulate the three key modules: Personalization Module, Keyword Extractor, and Response Generator.

Personalization Module takes in the input and feeds a processed input into keyword extractor. How to process the input can be customized depends on the task required. Personalization module exist as an interface for any user information system. We will explore methods to personalize responses with little user data(Su et al. 2019) as part of future work. Currently, the module post process the dialog history by parsing out personality descriptions.

Keyword Extractor takes in the input from personalization module and proposes a list of keywords for the user to choose. Moreover, the user can also change some of the propose keywords. The changed keyword list and user selection will be both passed to the keyword extractor and later response generator. As a result, the user provides keyword extractor a feedback loop to update the keyword suggestion.

Response Generator receives the input from keyword extractor in form of user keyword selection and generate a response using a pretrained language model such as Dialog GPT. The system will propose the candidate response to the user and ask for permission to speak. The user can either approve the response or make corresponding edits. The edit will be feed back into the response generator for the second time and compute the type of edits. Depends on the nature of the experiment, response generator can customize reward function based off the edit user makes. Thus, fine-tuning the pretrained language model with reinforcement learning algorithms.

Training Pipeline and Evaluation

We will use persona chat (Zhang et al. 2018) dataset to mimic real world settings during the conversation and demonstrate the user interface. Persona data provides a persona list for each conversation episode. When the training session starts, persona dataset teacher class in ParlAI will first feed the dialog history along with the persona into keyword extractor. For demonstration purposes, we used a fixed set of initial keywords. However, the keywords can be both edited, selected, and deleted. Figure 2 display the UI for keyword extraction.

In order to minimize user keystroke, user interface can operate with only left, right, and enter. The insight behind this design is original ACAT are specialized in facial tracking and minimal input detection for users with motor neuron disease. We also provide optional word-editing features. The interface feeds both user edited keywords and the selected keyword into the back-end system. The response generator, provided with the input from keyword extractor, uses a pretrained DialogGPT (Zhang et al. 2020) that we fine-tune the last layer to propose a candidate response. The candidate response is proposed to the user in Figure 3, where the user can make edits and respond with the edited output. Moreover, the user can edit the whole sentence if decide the proposed response is inadequate. Our system can review the changes made by the user and provide the response generator a reward depend on the amount of edits the user makes. Furthermore, the selections made by the user will be stored in playback sessions to allow the user to review the conversation out of the current dialog scope. This minimalist user interface adhered to our design goal on simplifying while maximizing the control, and provided a constant interactive system optimization.

Discussion

This framework is the first among a series of projects to improving assisted language generation system. By introducing human feedback into the training loop, we have provided a interface that can connect front-end design adapted to user's need and back-end existing offline dialog task training framework. By allowing the user to actively edit the response provided by the response generator, we have provided an interface for reinforcement learning experiment setup to fine tune language models.

While this project marks the progress on the overall framework, each individual components require more detailed design and expansion. Future plans include adding internal session for reply learning on personalization module, replay session to allow users to provide detailed feedback to the system, explicit design in user interface to optimize dialog flow, and experiment different reinforcement learning algorithms such as PPO and TRPO.

References

Gao, X.; Zhang, Y.; Galley, M.; Brockett, C.; and Dolan, B. 2020. Dialogue Response RankingTraining with Large-Scale Human Feedback Data. In *EMNLP*.

Gupta, P.; Kumar, V. B.; Bhutani, M.; and Black, A. W. 2019. WriterForcing: Generating more interesting story endings. *arXiv preprint arXiv:1907.08259* .

Miller, A.; Feng, W.; Fisch, A.; Lu, J.; Batra, D.; Bordes, A.; Parikh, D.; and Weston, J. 2017. ParlAI: A Dialog Research Software Platform. *arXiv preprint arXiv:1705.06476* .

Nachman, L.; Prasad, S.; et al. 2018. ACAT: Assistive Context Aware Toolkit - https://01.org/acat/. Date accessed: 2020-11-01.

Radford, A.; Narasimhan, K.; Salimans, T.; and Sutskever, I. 2018. Improving Language Understanding by Generative Pre-Training. URL https://openai.com/blog/language-unsupervised. Date accessed: 2020-11-01.

See, A.; Roller, S.; Kiela, D.; and Weston, J. 2019. What makes a good conversation? How controllable attributes affect human judgments. In *Proceedings of the 2019 NAACL Conference: Human Language Technologies, Volume 1*, 1702–1723.

Su, F.-G.; Hsu, A. R.; Tuan, Y.-L.; and Lee, H.-Y. 2019. Personalized Dialogue Response Generation Learned from Monologues. *Proc. Interspeech 2019* 4160–4164.

Welleck, S.; Kulikov, I.; Roller, S.; Dinan, E.; Cho, K.; and Weston, J. 2020. Neural Text Generation with Unlikelihood Training. *ArXiv* abs/1908.04319.

yang Wu, Q.; Li, L.; and Yu, Z. 2020. TextGAIL: Generative Adversarial Imitation Learning for Text Generation. *ArXiv* abs/2004.13796.

Zhang, S.; Dinan, E.; Urbanek, J.; Szlam, A.; Kiela, D.; and Weston, J. 2018. Personalizing Dialogue Agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2204–2213. Melbourne, Australia: Association for Computational Linguistics. doi:10.18653/v1/P18-1205. URL https://www.aclweb.org/anthology/P18-1205.

Zhang, Y.; Sun, S.; Galley, M.; Chen, Y.-C.; Brockett, C.; Gao, X.; Gao, J.; Liu, J.; and Dolan, B. 2020. DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation. In *ACL, system demonstration*.